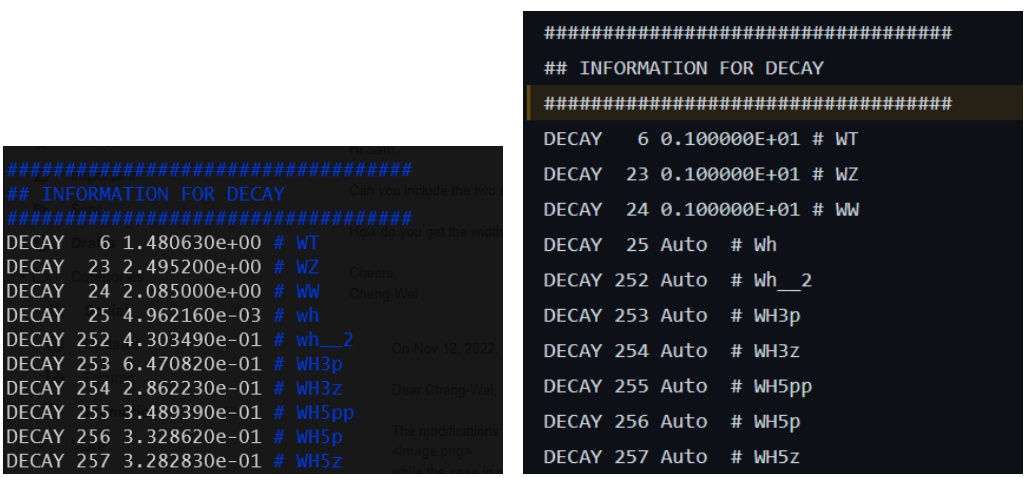
Comparison Notes

1 Preparation of sample

The samples are generated by the model `GM_UFO`. The model and parameters are the same as the paper. We want to comment that, according to Janice's code, she also modifies the value of the decay width of the top quark, W and Z boson, and all the exotic Higgs in the GM model when generating samples from MadGraph, even though she does not include this information in her paper. For completeness, we also take the decay width modification into account.

Figure 1 (a) shows the original decay widths in MadGraph. Figure 1 (b) is Janice's parameter card. She modified the decay width of the top quark, W, and Z boson to 1 GeV and exotic Higgses in the GM model to "Auto", meaning the decay widths are calculated by MadGraph.



(a) Original (b) Modified

Figure 1: Left is the original decay widths in MadGraph. The right is the decay widths in Janice's parameter card.

1.1 Sample selection

Sample selection criteria:

1. The transverse momentum of jets are required $p_{\mathrm{T}} \in (350, 450)$GeV and in range $|\eta| < 1$.
2. Merging: The angular distance between the two quarks decayed from the vector boson is required $\Delta R(q_1, q_2) < 0.6$.
3. Matching: The vector boson and jet are matched if $\Delta R(V, j) < 0.1$.

The training samples are constructed from the vector bosons that pass these criteria. All the vector bosons passing the criteria are included, even for the events with only one vector boson passing the criteria.

1.2 Jet image

After the sample selection, we can construct the jet image of the matching vector boson jets. The jet image is constructed after preprocessing procedure, including centralization, rotation, flipping, and pixelating.

For some events, the vector boson jet will cross the $\phi = \pm\pi$ boundary, which will make the centralization step fail. We will plus the ϕ coordinate of jet constituents by π so that its center is around $\phi = 0$. This has also have done in Janice's code.

1.3 Plots

For the following figures, the sample sizes are $(W^+, W^-, Z) = (350\text{ k}, 364\text{ k}, 316\text{ k})$.

Figure 2 demonstrates the jet mass distribution.

Figure 3 shows the jet charge distributions with different κ.

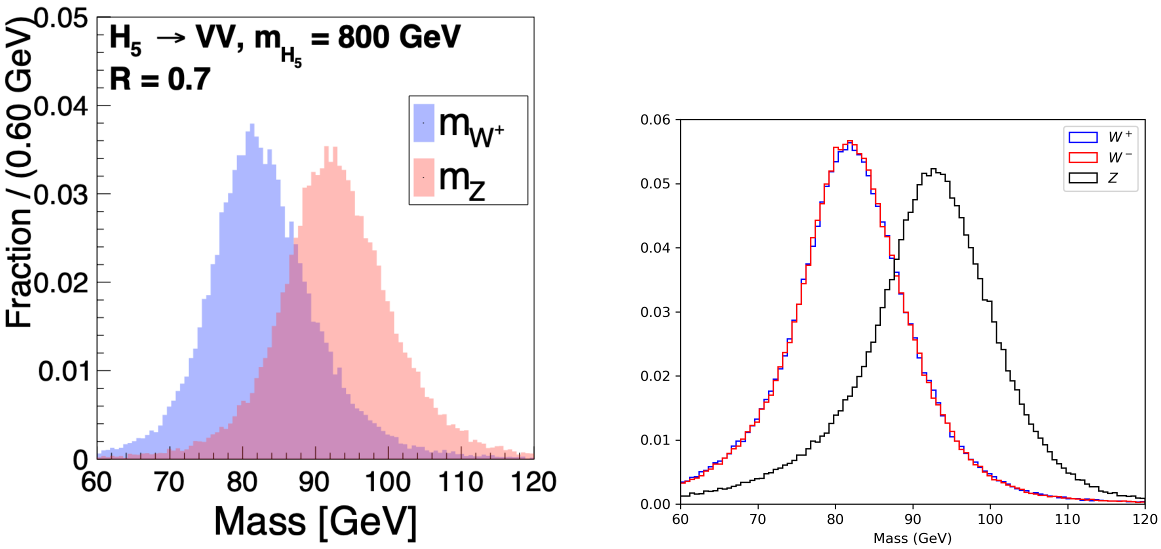
Figure 4, 5 shows the average jet images of p_{T} and $\mathcal{Q}_\kappa$.

Figure 6 presents the Z jet image minus W^+ jet image in p_{T} channel.

2 Ternary classification

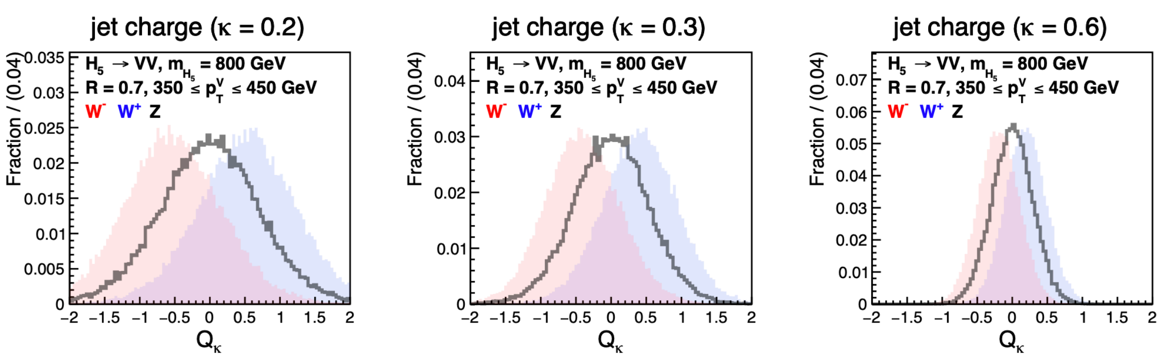
The training, validation, and testing sample sizes are shown in Table 1. The BDT training features are jet mass and jet charge. The CNN training features are the jet image with p_{T} and $\mathcal{Q}_\kappa$.

The training results are summarized in Table 2.

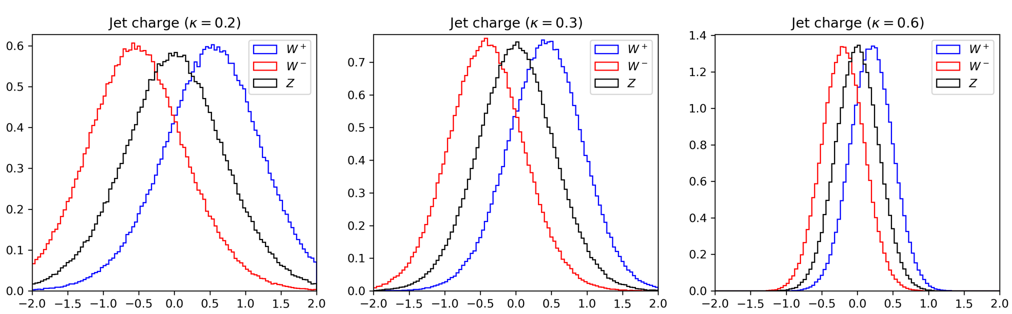


(a) Paper (b) Ours

Figure 2: Jet mass distribution of their vector boson samples.

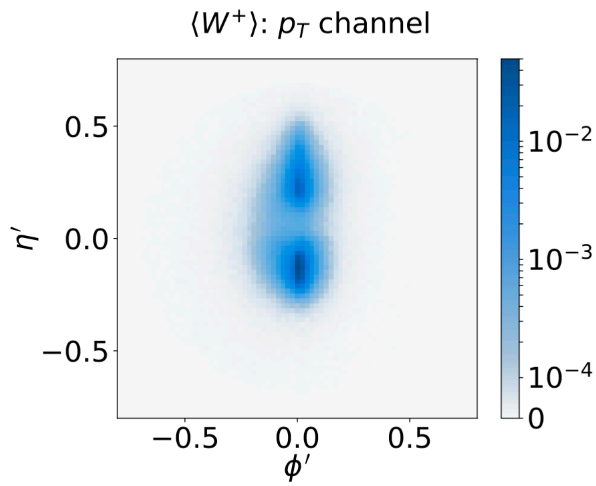


(a) Paper

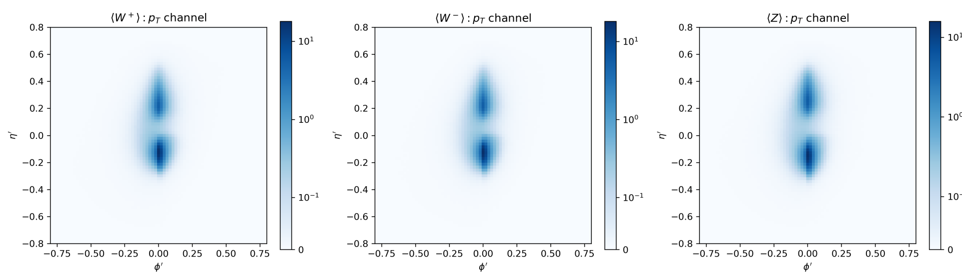


(b) Ours

Figure 3: $\mathcal{Q}_\kappa$ distribution of vector boson samples.



(a) Paper

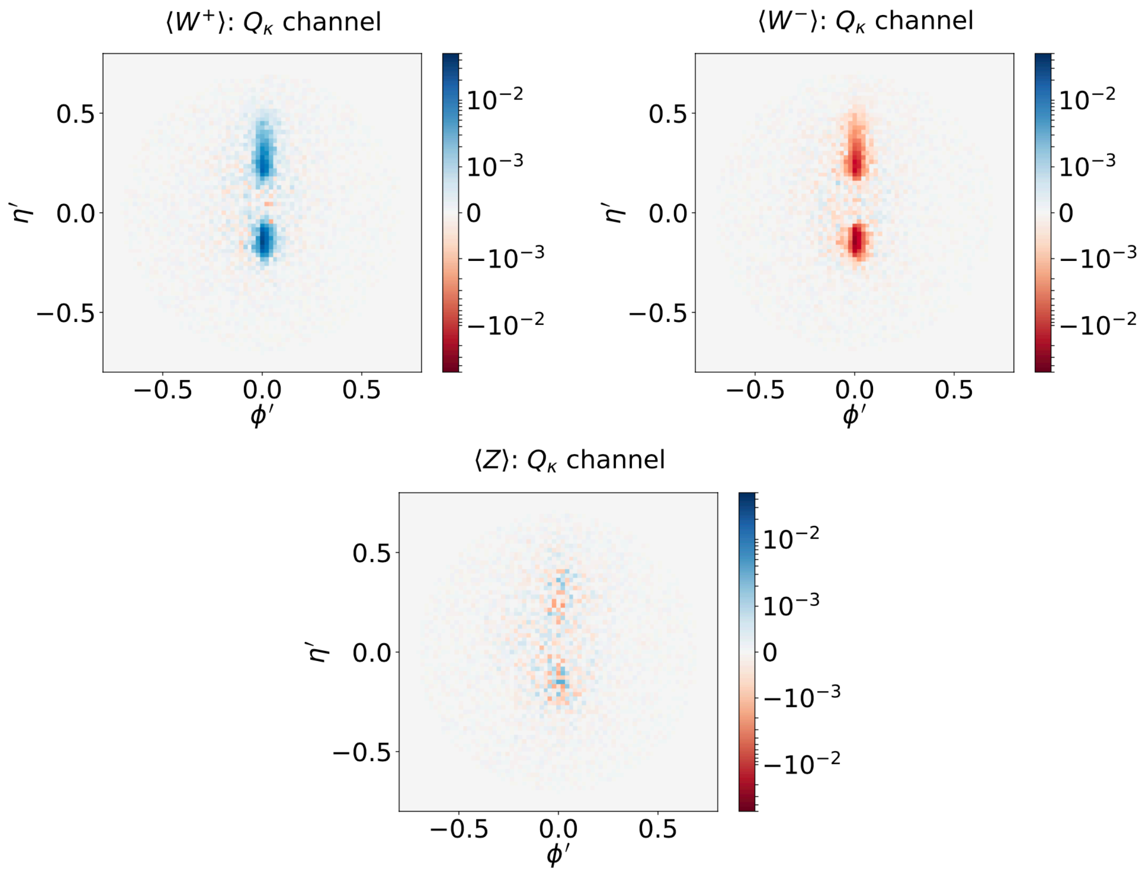


(b) Ours

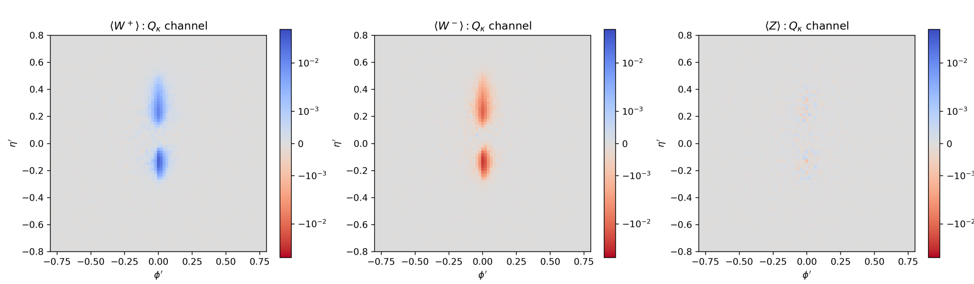
Figure 4: Average of jet images in the p_{T} channel.

Table 1: The entries in the sum correspond to the (W^+, W^-, Z) samples.

Case	Training set	Validation set	Testing set
Paper	$169k + 178k + 157k$	$19k + 20k + 18k$	$60k + 64k + 55k$
1	$224k + 232k + 201k$	$56k + 58k + 50k$	$69k + 72k + 63k$

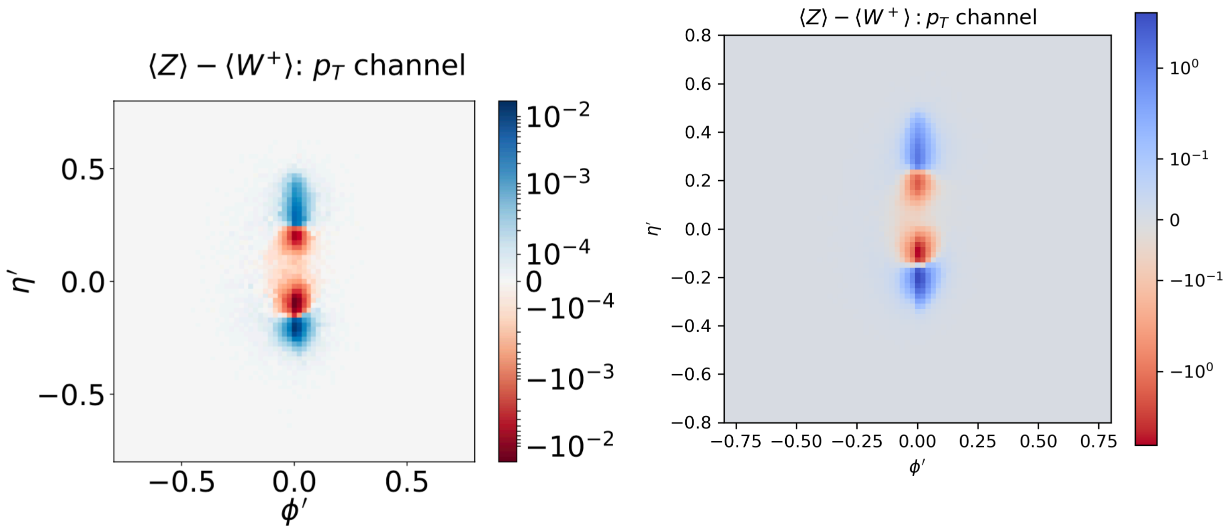


(a) Paper



(b) Ours

Figure 5: Average of jet images in the $\mathcal{Q}_\kappa$ channel, with $\kappa = 0.15$.



(a) Paper

(b) Ours

Figure 6: The difference between the Z and W^+ average jet images in p_T channel.

Table 2: The training results of Janice's paper and ours. The AUC and ACC values of the testing results, expressed in percentage, for BDT, CNN, and CNN^2 models. The CNN and CNN^2 models are presented with an average performance and a standard deviation value since they are trained over ten times with the same training data set. Yet, the value in the parenthesis for the ACC value of each boson is only derived from a single result.

			W^+		W^-		Z	
	Sample	Overall ACC	AUC	ACC	AUC	ACC	AUC	ACC
BDT $\kappa = 0.30$	Paper	66.8	N.A.	N.A.	83.5	75.8	83.7	77.3
CNN $\kappa = 0.15$		72.0	N.A.	N.A.	87.2	78.9	89.4	81.7
CNN^2 $\kappa = 0.15$		73.2	N.A.	N.A.	87.6	79.5	90.9	83.3
BDT $\kappa = 0.30$	1	65.9	82.9	77.7	82.9	77.5	80.9	78.4
CNN $\kappa = 0.15$		69.76±0.06	86.46±0.02	(80.4)	86.34±0.02	(79.6)	83.80±0.04	(81.2)
CNN^2 $\kappa = 0.15$		69.80±0.05	86.30±0.02	(80.0)	86.19±0.02	(79.7)	84.39±0.05	(81.8)

By doing the CNN and CNN^2 training many times, we found that the phase transition stated in Janice's paper seldom shows up. For most training processes, the accuracy ends up with 65% in the first epoch and smoothly increases to the highest value. The loss and learning curve of CNN and CNN^2 models are illustrated in Figure 7. We also show the ROC curve in Figure 8

Here we provide the loss and learning curve and the ROC curve for the CNN^2 model in Figure 9 and Figure 10.

3 Correct decay width sample

In this section, the sample is generated with the correct decay widths, i.e., the decay width of t, W, and Z do not change and the exotic Higgses in the GM model are set to "Auto", meaning the decay widths are calculated by MadGraph.

3.1 Training results

The training, validation, and testing sample size are in Table 3.

Table 3: The entries in the sum correspond to the (W^+, W^-, Z) samples.

Case	Training set	Validation set	Testing set
1	$223k + 232k + 202k$	$56k + 57k + 50k$	$69k + 72k + 62k$

The training results are summarized in Table 4. The results are worse than Sec.2. It

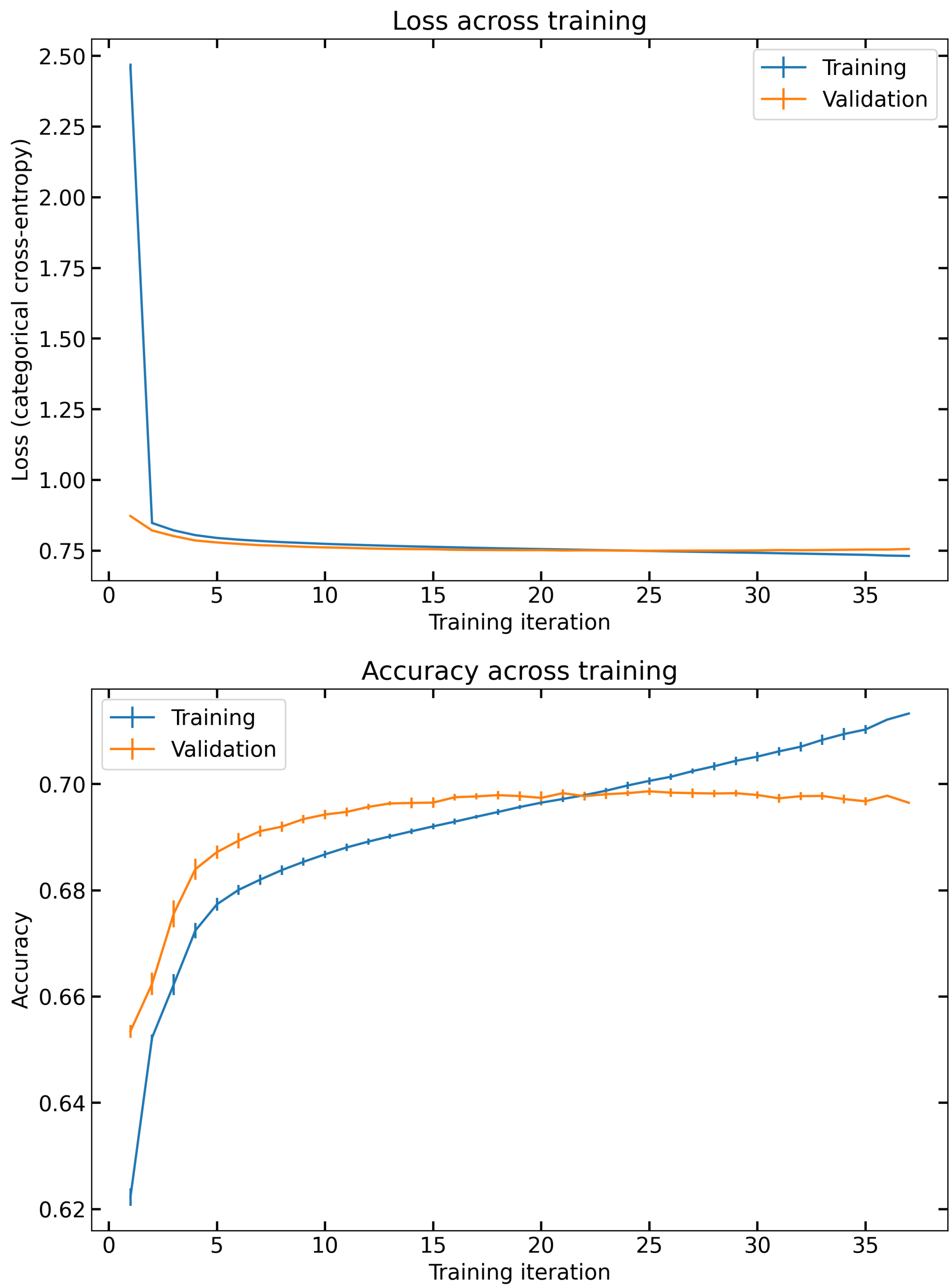
Table 4: The training results of correct decay width sample. The training results of CNN and CNN^2 are presented with an average and a standard deviation. These values are derived from 10 times training with the same dataset. Yet, the ACC value of each boson is only derived from a single result.

			W^+		W^-		Z	
	Sample	Overall ACC	AUC	ACC	AUC	ACC	AUC	ACC
BDT $\kappa = 0.30$		65.1	82.6	77.5	82.5	77.0	79.2	77.4
CNN $\kappa = 0.15$	1	68.24±0.04	85.72±0.02	(79.5)	85.58±0.02	(79.1)	81.64±0.05	(79.8)
CNN^2 $\kappa = 0.15$		68.26±0.13	85.55±0.04	(79.5)	85.41±0.05	(79.0)	82.22±0.08	(80.2)

seems that the decay width of t, W, and Z will affect the training results.

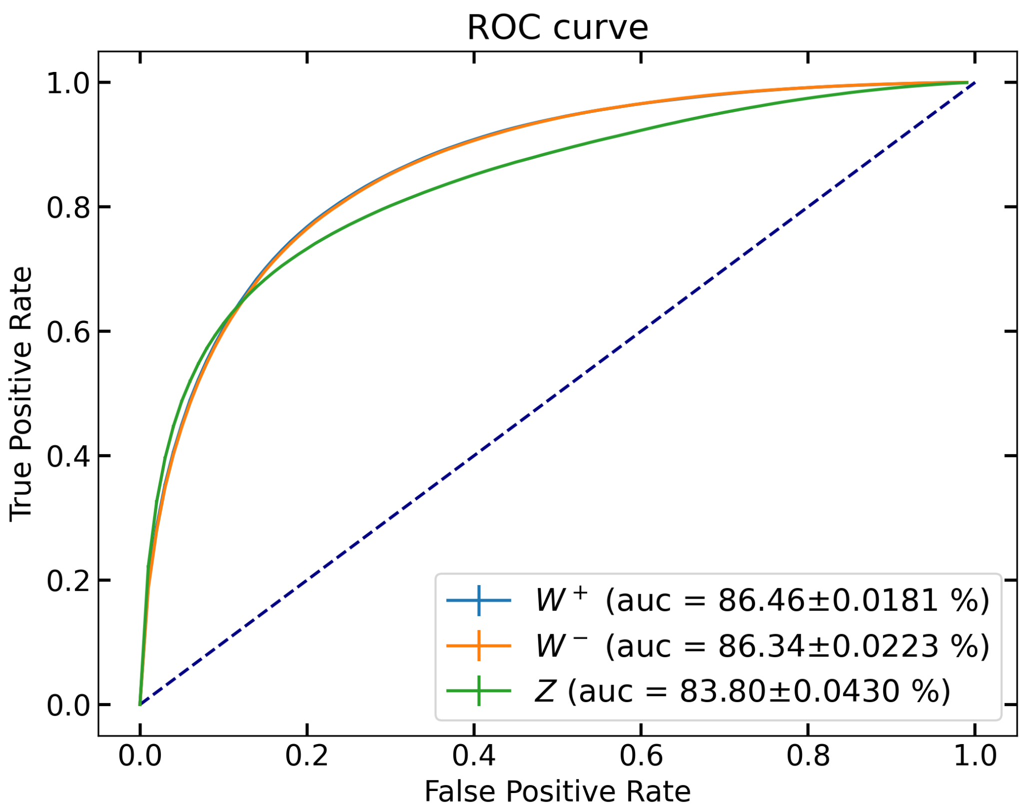
Figure 11 is CNN's loss and accuracy curve. Figure 12 is CNN's ROC curve.

Figure 13 is CNN^2's loss and accuracy curve. Figure 14 is CNN^2's ROC curve.



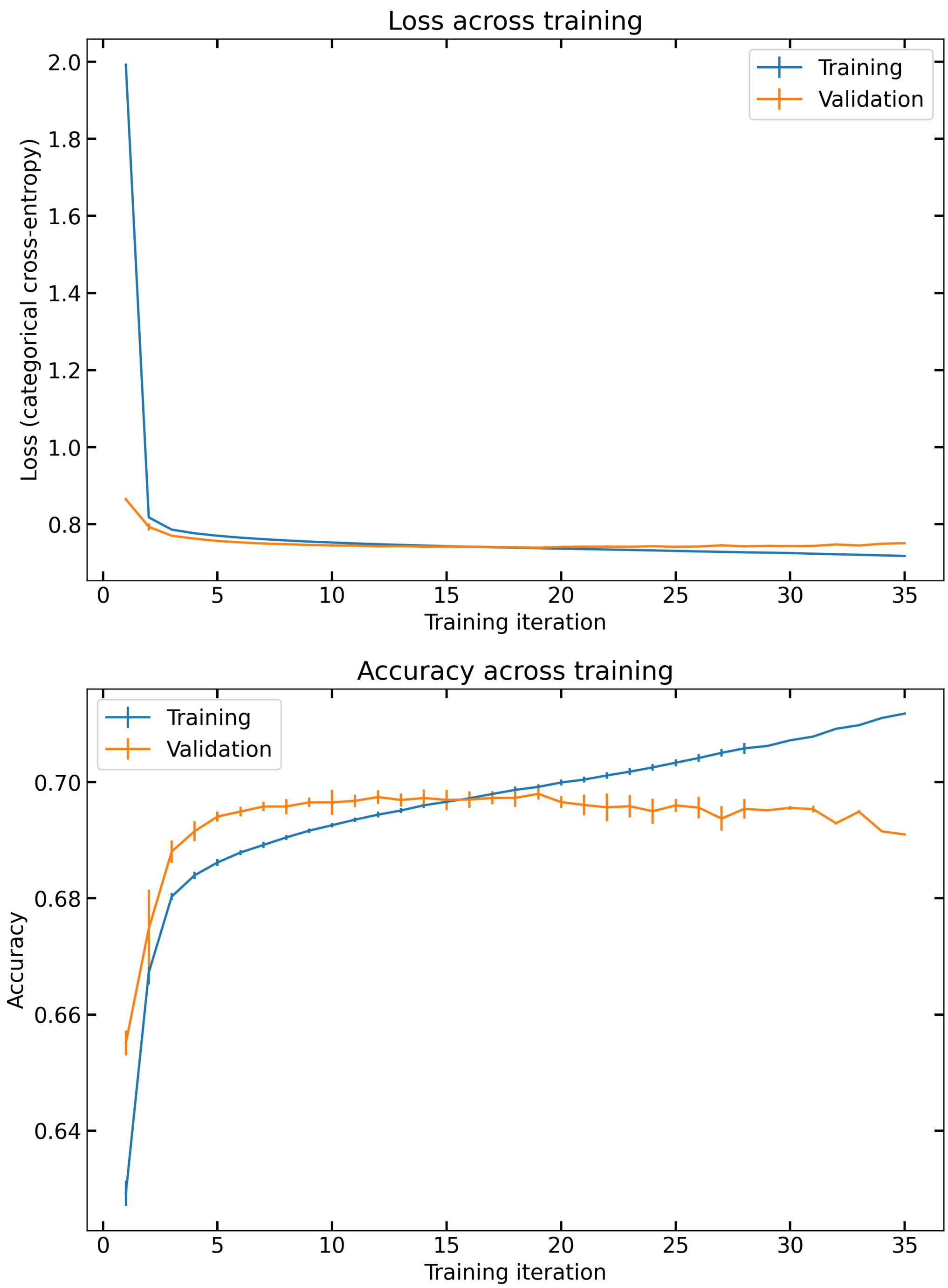
(a) Ours

Figure 7: The loss and learning curve for the CNN model. Both of them are demonstrated with the average value (solid curve) and the first standard deviation range (error bar).



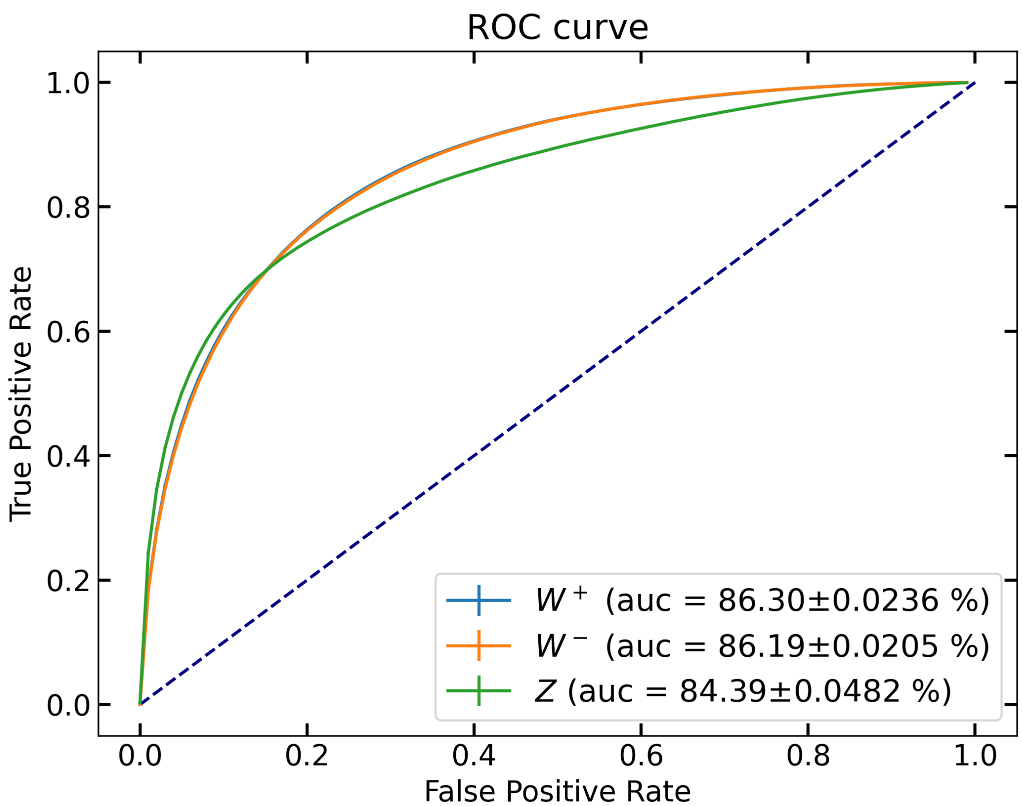
(a) Ours

Figure 8: The ROC curve of each vector boson for the CNN model. The plotting scheme is the same as Figure 7.



(a) Ours

Figure 9: The loss and learning curve for the CNN^2 model. The plotting scheme is the same as Figure 7.



(a) Ours

Figure 10: The ROC curve of each vector boson for the CNN^2 model. The plotting scheme is the same as Figure 7.

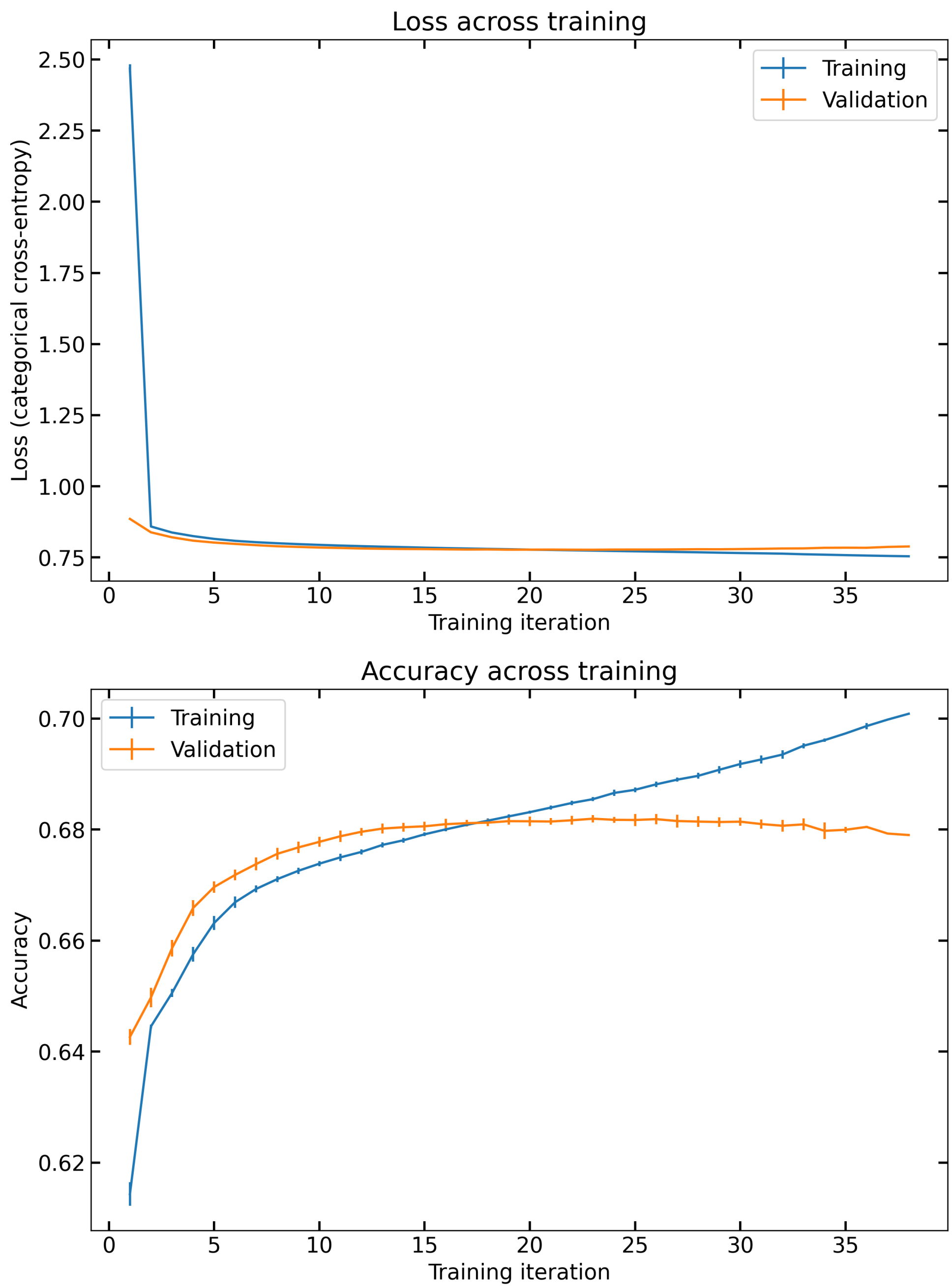


Figure 11: The loss and accuracy curve for the CNN model. Both of them are demonstrated with the average value (solid curve) and the first standard deviation range (error bar).

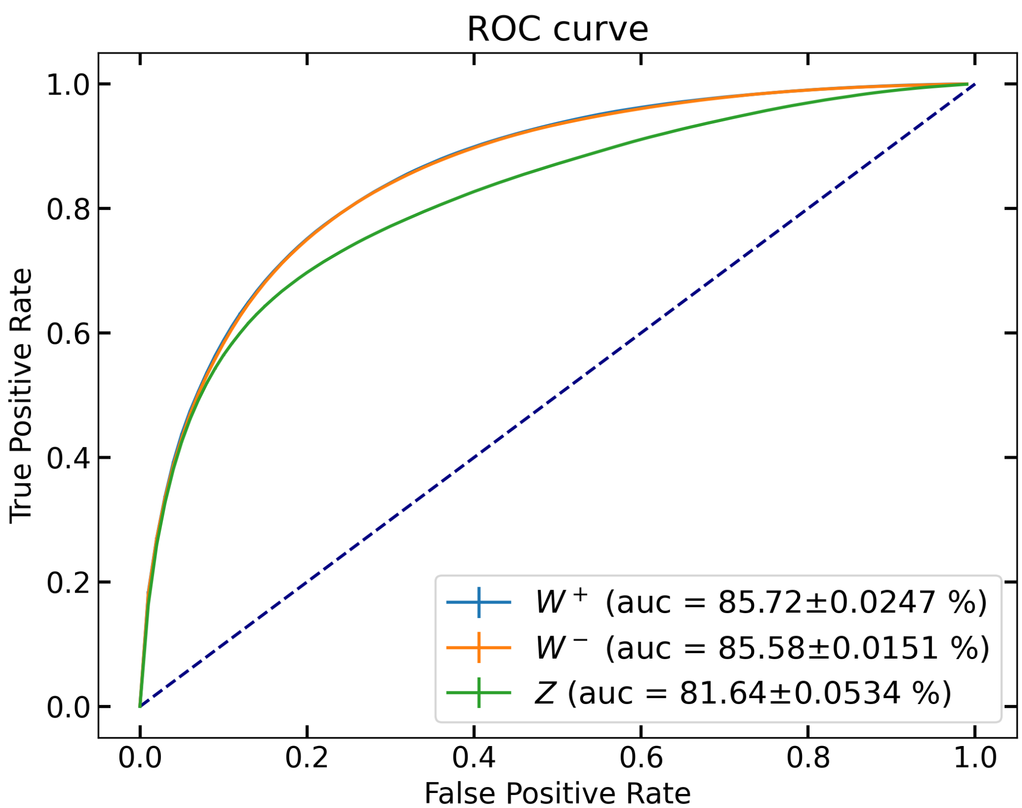


Figure 12: The ROC curve of each vector boson for the CNN model. The plotting scheme is the same as Figure 11.

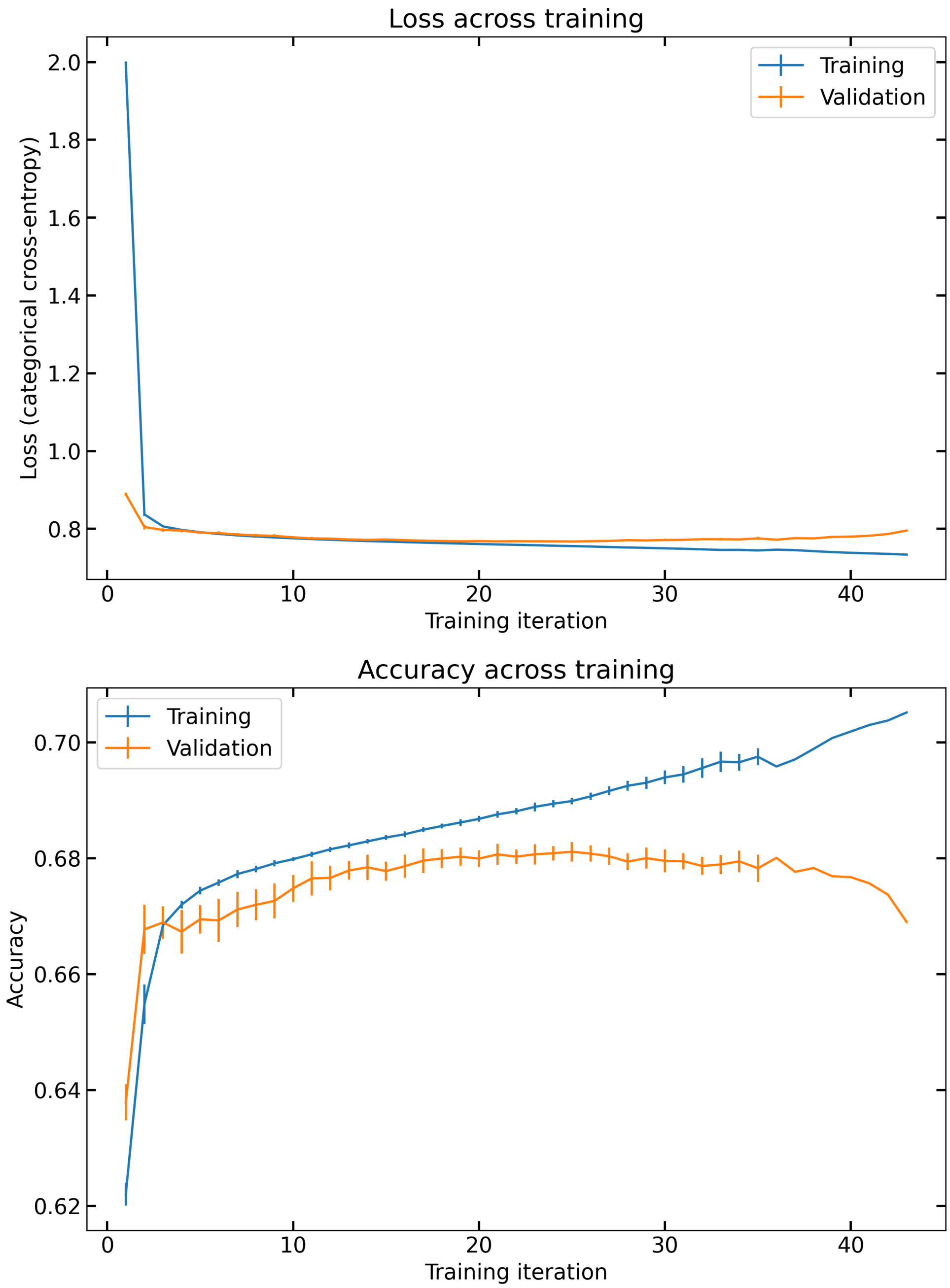


Figure 13: The loss and accuracy curve for the CNN^2 model. The plotting scheme is the same as Figure 11.

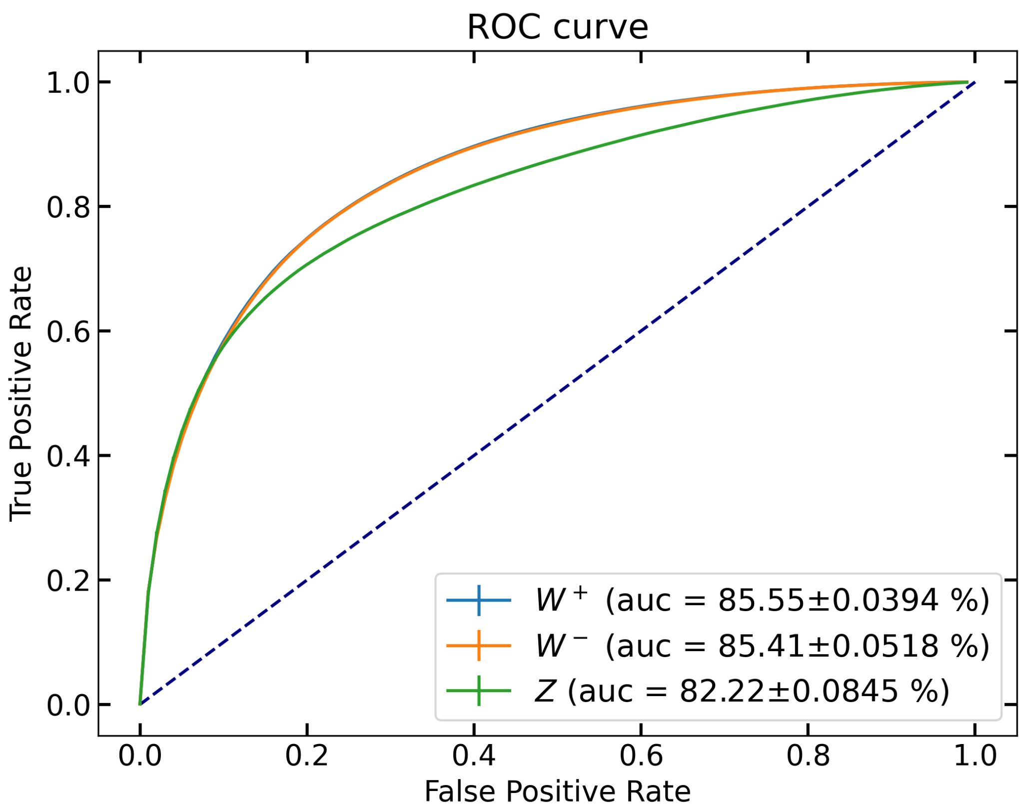


Figure 14: The ROC curve of each vector boson for the CNN^2 model. The plotting scheme is the same as Figure 11.